

Image reconstruction in optical interferometry: Benchmarking the regularization

S. Renard¹, E. Thiébaud², and F. Malbet¹

¹ IPAG, CNRS/UMR 5571, Université J. Fourier, BP-53, 38041 Grenoble Cedex, France
e-mail: Stephanie.Renard, Fabien.Malbet@obs.ujf-grenoble.fr

² Centre de Recherche Astrophysique de Lyon, CNRS/UMR5574, 69561 St-Genis-Laval, France
e-mail: thiebaut@obs.univ-lyon1.fr

Received; accepted

ABSTRACT

Context. With the advent of visible and infrared long-baseline interferometers with more than two telescopes, both the size and the completeness of interferometric data sets have significantly increased, allowing images based on models with no a priori assumptions to be reconstructed with an aperture synthesis technique.

Aims. Our main objective is to analyze the multiple parameters of the image reconstruction process with particular attention to the regularization term and the study of their behavior in different situations (types of astrophysical objects, telescope array configurations, level of noise, *etc.*). The secondary goal is to derive practical rules for the users.

Methods. Using the Multi-aperture image Reconstruction Algorithm (MiRA), we performed multiple systematic tests, analyzing 11 regularization terms commonly used. The tests are made on different astrophysical objects, different (u, v) plane coverages and several signal-to-noise ratios to determine the minimal configuration needed to reconstruct an image. We establish a methodology and we introduce the *mean-square errors* (MSE) to discuss the results.

Results. From the ~ 24000 simulations performed for the benchmarking of image reconstruction with MiRA, we are able to classify the different regularizations in the context of the observations. We find typical values of the regularization weight. A minimal (u, v) coverage is required to reconstruct an acceptable image, whereas no limits are found for the studied values of the signal-to-noise ratio. We also show that super-resolution can be achieved with increasing performance with the (u, v) coverage filling.

Conclusions. Using image reconstruction with a sufficient (u, v) coverage is shown to be reliable. The choice of the main parameters of the reconstruction is tightly constrained. We recommend that efforts to develop interferometric infrastructures should first concentrate on the number of telescopes to combine, and secondly on improving the accuracy and sensitivity of the arrays.

Key words. Instrumentation: interferometers – Techniques: interferometry – Techniques: image processing

1. Introduction

Many astrophysical studies require milli-arcsecond (hereafter mas) resolution images at optical wavelengths (visible and infrared), for example, the understanding of the interplay between accretion and ejection in the inner part of the disks of young stellar objects, the expansion mechanisms in novae just a few hours or days after the explosion, and the nature of dust in active galactic nuclei. Information at such a high resolution at the optical wavelengths requires diffraction-limited images with pupil sizes of the order of tens to hundreds of meters that can only be achieved by interferometrically combining light from separate apertures. The *Very Large Telescope Interferometer* (VLTI) and the *Center for High Angular Resolution Array* (CHARA) are facilities that provide interferometric measurements that can be used to reconstruct images of stellar surfaces (e.g., Monnier et al. 2007; Haubois et al. 2009; Zhao et al. 2009), binaries (e.g., Zhao et al. 2008; Kraus et al. 2009), circumstellar shells around evolved stars (e.g., Le Bouquin et al. 2009), and the close environment of young stars (Renard et al. 2010; Kraus et al. 2010).

Owing to the sparse (u, v) coverage, the image reconstruction process is ill posed as there are more unknowns, e.g. the pixels of the image, than measurements. The data alone are insufficient to reconstruct an unambiguous image and some additional con-

straints, so-called the regularizations, are needed to converge to a unique and stable solution. Compared to radio interferometry, the data are much sparser in optical interferometry; hence, we expect the image reconstruction problem to be much more sensitive to the choice and the tuning of the regularization. Since the general study of regularization by Titterton (1985), many different methods have been proposed in the literature to adjust the regularization level. Since image reconstruction for optical interferometry is still in its infancy, it is fundamental to analyze the different types of regularization to find those that are the most suitable for the different astrophysical problems and to be able to tune the weight of the regularization. In this context, we carried out systematic tests using the image reconstruction algorithm devoted to optical interferometry data developed by Thiébaud (2008), called the Multi-aperture image Reconstruction Algorithm (MiRA). The analysis of these tests allow us to extract some general conclusions and establish practical rules for the users.

The mathematical principles of the image reconstruction technique is presented in Sect. 2. The parameters of the simulated data are presented together with the characteristics of the images and the strategy in Sect. 3. The results of the simulations are presented and discussed in Sect. 4, with an analysis of the role of the different terms and parameters in the image reconstruction. Finally, our conclusions are summarized in Sect. 5.

2. Principles of image reconstruction from optical interferometric data

We do not intend to provide a formal and precise description of image reconstruction in optical interferometry, but instead sufficient details to ensure that the paper is self-contained. Readers interested by a more detailed description of the method should refer to [Thiébaud & Giovannelli \(2010\)](#).

2.1. Data from optical interferometric observations

The principle of interferometry is to recombine coherently the beams from two or more independent telescopes and measure the so-called complex visibilities of the fringe patterns produced by the interferences. According to the van Cittert-Zernicke theorem, for an ideal interferometer, the complex visibility $V_{j_1, j_2}(t)$ of the fringes produced by the interferences of the telescopes j_1 and j_2 at time t is proportional to the Fourier transform of the object brightness distribution $\hat{I}(\mathbf{v}_{j_1, j_2}(t))$ at spatial frequency $\mathbf{v}_{j_1, j_2}(t) = \mathbf{B}_{j_1, j_2}^\perp(t)/\lambda$, where λ is the wavelength, and the so-called *baseline* $\mathbf{B}_{j_1, j_2}^\perp(t)$ represents the separation between the two telescopes projected on a plane perpendicular to the line of sight ([Lawson 2000](#); [Malbet & Perrin 2007](#)). Since the number of measurements is finite, to simplify the equations we introduce some notation for the m -th measured complex visibility and the corresponding spatial frequency, given by

$$V_m \stackrel{\text{def}}{=} V_{j_{1,m}, j_{2,m}}(t_m), \quad (1)$$

$$\mathbf{v}_m \stackrel{\text{def}}{=} \mathbf{B}_{j_{1,m}, j_{2,m}}^\perp(t_m)/\lambda, \quad (2)$$

where $j_{1,m}$ and $j_{2,m}$ are the interfering telescopes and t_m is the time of observation.

2.2. Description of the image model

The final product of the image reconstruction is an image that can be treated as a grid of square pixels. In this context, the object brightness distribution as a function of the position θ can be approximated using the parametrization

$$I(\theta) = \sum_{n=1}^N x_n b_n(\theta), \quad (3)$$

where $\mathbf{x} = \{x_n\}_{n=1}^N$ are the image parameters, *e.g.* the pixel values of the image, and $\{b_n(\theta)\}_{n=1}^N$ is the chosen basis of functions, *e.g.* the response function of each pixel. The image reconstruction then consists of estimating the N parameters \mathbf{x} that most closely fit the interferometric data. In this paper, we chose x_n to be proportional to the value of the n -th pixel of a sampled image and $b_n(\theta) = b(\theta - \theta_n)$, where $b(\theta)$ is the pixel shape and θ_n the position of the n -th pixel; thus

$$x_n \stackrel{\text{def}}{=} \alpha I(\theta_n), \quad (4)$$

where $\alpha > 0$ is a scaling factor such that \mathbf{x} is normalized (this is required by the interferometric data format, *cf.* [Pauls et al. \(2005\)](#)). With this model, the exact Fourier transform of the brightness distribution is given by

$$\hat{I}(\mathbf{v}) = \sum_n x_n \hat{b}_n(\mathbf{v}) = \hat{b}(\mathbf{v}) \sum_n x_n e^{-i2\pi \theta_n \cdot \mathbf{v}}, \quad (5)$$

where $\hat{b}_n(\mathbf{v})$ and $\hat{b}(\mathbf{v})$ are the Fourier transforms of the basis functions. In our case, they correspond to the Fourier transform of the

pixel response function, *i.e.* the pixel shape. Hence, the model of the m -th complex visibility is given by

$$\hat{I}_m = \hat{I}(\mathbf{v}_m) = \sum_n A_{m,n} x_n = (\mathbf{A} \cdot \mathbf{x})_m, \quad (6)$$

where \mathbf{A} is a matrix with the complex coefficients

$$A_{m,n} = \hat{b}_n(\mathbf{v}_m) = \hat{b}(\mathbf{v}_m) e^{-i2\pi \theta_n \cdot \mathbf{v}_m}. \quad (7)$$

This matrix multiplication performs a linear transformation that contains the Fourier transform, the pixel shape, and the sparse sampling of the (u, v) plane.

The main problem in optical interferometry is the small number of telescopes (currently up to four or six), which leads to a sparse sampling of the spatial frequencies, the so-called (u, v) plane (see Fig. 2). Owing to random effects caused by the atmospheric turbulence, the visibility phase cannot be calibrated and the power spectrum and the closure phase are used. This results in a partial loss of the Fourier phase information ([Thiébaud & Giovannelli 2010](#)).

2.3. Inverting the problem of interferometric imaging

Because of the sparse (u, v) coverage and the possible lack of other information such as the phase, the reconstruction of an image obtained by the interferometric data alone is an ill-posed inverse problem. It needs additional *a priori* constraints to be recasted into a problem that has a unique and stable solution. A general prescription is to express the solution as one that minimizes a penalty function f under some strict constraints ([Thiébaud 2005](#); [Thiébaud & Giovannelli 2010](#))

$$\mathbf{x}^+ = \arg \min_{\mathbf{x}} f(\mathbf{x}) \quad \text{s.t.} \quad \mathbf{x} \geq 0 \text{ and } \sum_n x_n = 1, \quad (8)$$

with

$$f(\mathbf{x}) = f_{\text{data}}(\mathbf{x}) + \mu f_{\text{prior}}(\mathbf{x}), \quad (9)$$

where the so-called *likelihood* term $f_{\text{data}}(\mathbf{x})$ measures the discrepancy between the model and the available data, while the so-called *regularization* term $f_{\text{prior}}(\mathbf{x})$ measures the discrepancy with the *prior* information. In other words, minimizing the likelihood term $f_{\text{data}}(\mathbf{x})$ enforces the fit with the actual data, while minimizing the regularization term $f_{\text{prior}}(\mathbf{x})$ enforces the agreement with the priors. The so-called hyperparameter $\mu > 0$ is used to adjust the relative weight of the constraints set by the measurements and the ones set by the priors. In Eq. (8), the positivity ($\mathbf{x} \geq 0$) and the normalization ($\sum_n x_n = 1$) of the brightness distribution are also included by default.

For the image reconstruction, we use MiRA ([Thiébaud 2008](#)) to find solutions of Eq. (8). MiRA can deal with the various kinds of data provided by an optical interferometer and implements a number of different regularizations ([Thiébaud 2008](#); [Thiébaud & Giovannelli 2010](#)).

2.3.1. The likelihood term f_{data} and the data model

We focus on the choice of the priors and the tuning of their parameters. It is not within the scope of the paper to deal with global optimization issues and the search for a global minimum of the penalty function in Eq. (9). We therefore assume that the available measurements consist of complex visibilities, *i.e.* amplitude and phase, in order to have a *convex* likelihood term $f_{\text{data}}(\mathbf{x})$. If the regularization term $f_{\text{prior}}(\mathbf{x})$ is also convex, the global penalty function $f(\mathbf{x})$ will be convex, which ensures that

the solution of Eq. (8) is unique. Current optical interferometers only provide phase closures and power-spectrum data (*i.e.* the phase of the bispectrum and the squared amplitude of the complex visibilities), this means that our assumption will give somewhat optimistic results because some Fourier phase information is missing and because the likelihood term $f_{\text{data}}(\mathbf{x})$ is non-convex when dealing with real data. However, as the number of simultaneous interfering telescopes increases, the number of missing phases becomes much less important and they can be reliably derived using *self-calibration* (Pearson & Readhead 1984) to achieve a situation similar to the case studied in our simulations. Moreover, new interferometers will make use of a phase reference source to directly measure the phase of the complex visibilities (Delplancke et al. 2003).

However, the OI-FITS standard (Pauls et al. 2005) imposes the use of complex visibilities in their polar representation with independent error bars. We therefore simulate each measured complex visibility as

$$\varrho_m = |V_m| + \delta\varrho_m, \quad \text{and} \quad \varphi_m = \arg V_m + \delta\varphi_m, \quad (10)$$

where ϱ_m and φ_m are the measured amplitude and phase of the m -th measure, V_m the corresponding complex visibility computed from the true object brightness distribution, and $\delta\varrho_m$ and $\delta\varphi_m$ are additive noise terms. In our simulations, the noise terms have independent Gaussian statistics such that

$$\text{Var}(\delta\varrho_m) = \langle \varrho_m \rangle^2 \text{Var}(\delta\varphi_m), \quad (11)$$

where $\langle \varrho_m \rangle = |V_m|$ is the expected value of the amplitude that can be computed from the complex visibility of the true image. This particular choice follows approximately the model of Goodman (1985).

In this context, to define the likelihood term we use the *local approximation* (Meimon et al. 2005)

$$f_{\text{data}}(\mathbf{x}) = \sum_m \left\{ \left(\frac{r_{\parallel,m}(\mathbf{x})}{\sigma_{\parallel,m}} \right)^2 + \left(\frac{r_{\perp,m}(\mathbf{x})}{\sigma_{\perp,m}} \right)^2 \right\}, \quad (12)$$

where $r_{\parallel,m}(\mathbf{x})$ and $r_{\perp,m}(\mathbf{x})$ are the two components of the complex residuals, $r_m(\mathbf{x}) = \varrho_m e^{i\varphi_m} - (\mathbf{A} \cdot \mathbf{x})_m$, respectively, along and orthogonally to the measured complex visibility. Given the error bars $\sigma_{\varrho,m}$ and $\sigma_{\varphi,m}$ of the amplitude and the phase of the complex visibility, the standard deviations in the components of the complex residuals are (Pauls et al. 2005)

$$\sigma_{\parallel,m} = \sigma_{\varrho,m} \quad \text{and} \quad \sigma_{\perp,m} = \varrho_m \sigma_{\varphi,m}. \quad (13)$$

2.3.2. The regularization term f_{prior}

In our simulations, we test 11 different regularization terms that are commonly used in image reconstruction methods and are implemented in MiRA (see Appendix A for detailed expressions).

1. **Quadratic smoothness**, which most closely describes a smooth image and helps us to avoid unmeasured high frequencies.
- 2-3. **Compactness**, which describes compactness in the image plane and hence smoothness in the Fourier plane (Le Besnerais et al. 2008). Two different cases were studied in the simulations, with penalties of the second and third orders with respect to the distance of the center of the field of view.

4. **Total variation** (hereafter TV), which minimizes the total gradient of the image and helps us to describe uniform areas in the sought image with steep but localized changes (Strong & Chan 2003).
5. ℓ_1 **smoothness**, which is useful for an extended object with sharp edges since it is linear for strong gradients.
- 6-8. ℓ_p -**norm** with $p = 1.5$, $p = 2$ and $p = 3$. For $p > 1$, the ℓ_p -norm regularization tends to produce a smooth image as it reduces the variance in the pixels.
- 9-11. **Maximum entropy methods** (hereafter MEM) aims to identify the least informative image consistent with the data (Gull & Skilling 1984; Narayan & Nityananda 1986). We try three types of entropies *MEM-sqrt*, *MEM-log*, and *MEM-prior*, respectively. The first two tend to reproduce an image with a flux spread across a minimum number of pixels. The last one is minimum when the image is as close as possible to a prior image. This prior image is the Gaussian that most closely reproduces the amplitude visibility data.

The different regularization terms are expected to behave as follows. The positivity and the normalization imposed in all the reconstructions are an ℓ_1 norm and lead to the sparsity of the solution, *i.e.* to a minimum number of bright pixels to explain the data. As most astrophysical images are smooth and compact, we expect that the regularizations that can describe these images will behave well, *i.e.* smoothness (quadratic or ℓ_1), compactness, and TV. The ℓ_p -norm regularization with $p = 2$ (and by extension for $p > 1$ as the regularization has the same behavior) has the tendency to force to zero the spatial frequencies that have not been measured (according to the Parseval theorem). Since the regularization has to interpolate correctly between the data, which is closer to the reality, it is not expected to give good results. Finally, the MEM-prior is expected to yield more reliable results than the other type of entropy because our choice of the *a priori* image can closely describe a compact object.

3. Description of the simulations

We now describe all the various simulated data that we compute for different objects, (u, v) coverages, and signal-to-noise ratios (SNR), as well as the parameters used for the image reconstructions.

3.1. Simulated data

Our simulated data sets are saved as OI-Fits files (Pauls et al. 2005) and depend on several setting of the object type, the (u, v) coverage, and the SNR. The 90 data files are available on the JMMC website¹.

3.1.1. Astrophysical objects

For our simulations, we consider ten astrophysical objects (see Fig. 1) that differ in term of their morphology and the typical length scales of their structures.

1. **LkH α** : the model of LkH α describes a compact object with a peak of intensity and a smooth envelope. The model comes from the 2004 International Imaging Beauty Contest organized by P. Lawson for the IAU (Lawson et al. 2004).

¹ <http://apps.jmmc.fr/oidata/shared/srenard/>

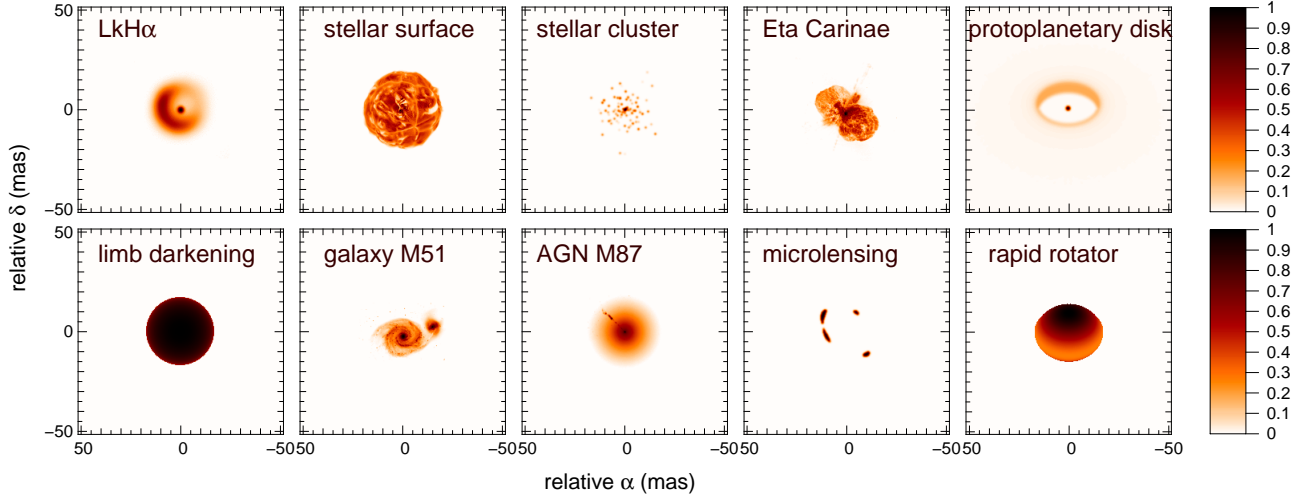


Fig. 1. Astrophysical objects used in the simulations.

2. **A stellar surface:** this is a model of the supergiant α Ori that presents some convective cells, producing small scales on a smooth background. This model was produced by A. Chiavassa for the 2009 Workshop on Interferometry Imaging (WII09) organized by J.-P. Berger and F. Malbet (Berger et al., in prep.).
3. **A stellar cluster:** the model consists of a hundred point sources. The position and the brightness of the sources follow a normal law.
4. **Eta Carinae:** the image of Eta Carinae presents many different scales and structures, such as the extended gas and the stars. This image was retrieved from the Hubble Space Telescope's website². Some treatments have been applied to the image, *i.e.* a mean of the three different color channels to produce a grayscale image and a cut of the low intensities to produce a null background.
5. **A protoplanetary disk:** the model represents an Herbig Ae/Be star with a point source (the star) and an extended structure (the disk). This model was computed by J.-P. Berger for the phase A science case of the VLTI-Spectro Imager instrument (Filho et al. 2008).
6. **A limb-darkened star:** we used the power-law model of Hestroffer (1997) with an exponent $\alpha = 0.3$. The image has a very smooth core with steep edges.
7. **The galaxy system M51:** this image of M51 consists of as many different scales and structures as Eta Carinae (gas, stars, spiral arms). This image was retrieved from the Hubble Space Telescope's website and was processed in a way similar to the image of Eta Carinae.
8. **The AGN M87:** this AGN has a jet that consists of a narrow structure surrounded by a smooth background due to the gas. The image was retrieved from the Hubble Space Telescope's website and the same treatments as the Eta Carinae's image have been applied.
9. **A gravitational microlensing image:** Gravitational microlensing is an astronomical phenomenon due to the gravitational lens effect. When a distant star or quasar becomes sufficiently aligned with a massive compact foreground object, the bending of light due to its gravitational field leads to two distorted unresolved images resulting in an observable magnification. The image shows four very compact structures. This model was developed by J. Surdej for the phase A

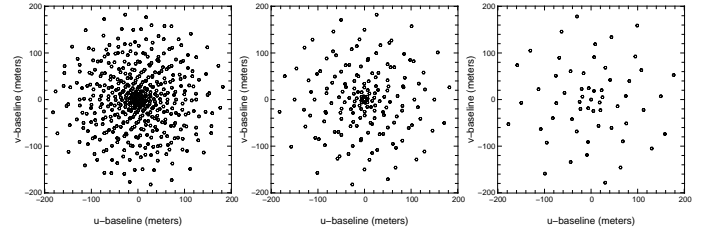


Fig. 2. (u, v) coverage. *From left to right:* rich (245 sampled frequencies), medium (88 sampled frequencies), and poor (31 sampled frequencies).

science case of the VLTI-Spectro Imager instrument (Filho et al. 2008).

10. **A rapid rotator:** the rapid rotation of a star affects the stellar shape and the local emitted flux. We used the model of Domiciano de Souza et al. (2002), with parameters $D = 0.78$ and $T_p = 35000$ K. The resulting light distribution was projected onto a plane with an inclination of 45° .

Since the goal of our tests is to determine the influence of the object's type on the image reconstruction, all the considered objects have a similar angular size of ~ 34 mas, which is consistent with the typical size of the field of view of an optical interferometer such as Amber/VLTI. To generalize our results, we expect the most important figure for a given object and instrument to be the number of *resolved elements*, which is equal to the ratio of the angular size of the object support to the effective resolution of the imaging system. Estimated by this ratio, the complexity of the objects that we have considered is in the range of 200 – 600 resolved elements.

3.1.2. (u, v) coverage

To study the influence of the instrumental configuration and analyze the capability of the regularization to interpolate the available data and thus fill the voids in the (u, v) plane, we consider several sparse (u, v) coverages (see Fig. 2). To remain general, our different (u, v) coverages were constructed to be uniformly spread. Each (u, v) coverage consists of concentric rings modulated by a sinusoid along the ring and with phases of the modulations chosen to maximize the distance between the points of two adjacent rings. This also avoids symmetries in the (u, v) cov-

² <http://hubblesite.org/>

erage. The concentration of the rings is more important on small baselines than on the largest ones to insure a good sampling of low spatial frequencies. In this paper, we consider three (u, v) coverages which differ in terms of the number of samples, called hereafter *rich* (245 sampled frequencies), *medium* (88 sampled frequencies), and *poor* (31 sampled frequencies) coverages. The chosen (u, v) configurations could be considered as very good by the standards of existing data, but the goal of the paper is not to cover all possible cases but to show the general trend.

3.1.3. Signal-to-noise ratio

To investigate the effects of varying the SNR, we use a SNR factor γ in the standard deviations given by

$$\sigma_{\varrho,m} = \gamma \langle \varrho_m \rangle \quad \text{and} \quad \sigma_{\varphi,m} = \gamma. \quad (14)$$

Therefore with these settings, the error bars become

$$\sigma_{\parallel,m} = \sigma_{\perp,m} = \gamma \langle \varrho_m \rangle. \quad (15)$$

To analyze the influence of the SNR on the reconstructed images, three values of γ are tested: *high* SNR (1%), *intermediate* SNR (5%), and *low* SNR (10%). We limit our study to these SNR values to maintain a small amount of results. Therefore, there is no systematic attempt to search for the limit of SNR and we realize that our worst case (10%) might be considered moderate noise in real data.

3.2. Parameters of the synthesized image

The resolution and the field of view (hereafter FOV) of the reconstructed image have to be chosen with care. On the one hand, the pixel size $\Delta\theta$ must be small enough to properly account for the highest spatial frequencies available in the data. Using Shannon's rule, we find that $\Delta\theta \leq \lambda/(2B_{\max})$, where B_{\max} is the maximum length of the observed baselines and λ the wavelength. In our reconstructions, the pixel size is one third of the limit set by Shannon's rule. This oversampling allows us to check whether image reconstruction can achieve some level of super-resolution. For instance, this yields $\Delta\theta = 0.4$ mas at $\lambda = 2.2 \mu\text{m}$ with $B_{\max} = 190$ m. On the other hand, the FOV has to be large enough to avoid field aliasing and therefore we take an image three times larger than the object itself. As all our objects have a size of 34 mas and thus approximately fit into 85×85 pixels, the reconstructed images have 256×256 pixels.

3.3. Reconstruction strategy

Given the data (determined by the object, the (u, v) coverage, and the noise realization) and the regularization, we conduct a sequence of image reconstructions for different values of the regularization weight μ . Since the problems solved are convex (as explained in Sect. 2.3.1), their solutions do not depend on the starting point. To reduce the calculation time, we therefore try to use a starting point that is as close as possible to the final solution. We begin the sequences of reconstructions with the highest value of μ and use the true image as the starting point for this first reconstruction. We then reduce μ and use the image previously obtained as the starting point. This latter step is repeated until we reach the lowest value of μ . Each reconstruction is an iterative process that is stopped when the norm of the gradient of the penalty function $f(\mathbf{x})$ is below a preset threshold. This threshold

is derived from the true image according to

$$\|\nabla f(\mathbf{x}^{\text{rec}})\| \leq \eta \frac{|f_{\mu}(\mathbf{x}^{\text{ref}})|}{\|\mathbf{x}^{\text{ref}}\|} = 10^{-5} |f_{\mu}(\mathbf{x}^{\text{ref}})|, \quad (16)$$

where $\eta > 0$ is a small value and $f_{\mu}(\mathbf{x}^{\text{ref}})$ the penalty computed for the true image and for each value of μ . The ℓ_2 norm of the true image is normalized and we assume that $\eta = 10^{-5}$.

3.4. Image quality criterion: the mean-squared error

To assess the quality of the reconstructed images, we consider the mean-squared error (hereafter MSE) of the reconstructed images. In our simulations, we use normalized input images with the same pixel size as for the reconstructed image. Hence, to compare the reconstructed image \mathbf{x}^{rec} and the true image \mathbf{x}^{ref} , we can simply use the MSE defined as

$$\text{MSE} = \frac{1}{N} \sum_n (x_n^{\text{rec}} - x_n^{\text{ref}})^2 = \frac{1}{N} \|\mathbf{x}^{\text{rec}} - \mathbf{x}^{\text{ref}}\|^2, \quad (17)$$

where N is the total number of pixels. We note that, since we use complex visibilities and our priors, apart from the FOV one, are shift-invariant, the reconstructed image is correctly centered and there is no need to compensate for registration errors.

4. Results and discussion

We performed a total of ~ 24000 simulations corresponding to the reconstruction of all cases described in Sect. 3 with different values of the hyperparameter μ . We present here the results of these simulations and analyze the consequences for the image reconstruction process in order to draw general conclusions.

We begin by discussing the optimal value of the hyperparameter and determining whether the MSE is a good quality criterion. We then discuss the effects of the following parameters:

- the regularization: what are the good and bad regularizations?
- the limits: what are the minimal (u, v) coverage and SNR value?
- the hyperparameter μ : what is the optimal value? With which parameters does it vary?
- the likelihood term: how does it vary? Can it be used to tune the regularization term instead of the hyperparameter μ ?
- the effective resolution: what degree of super-resolution can be achieved? How does it vary?

4.1. Optimal regularization weight μ^+

We first investigate whether there is an optimal regularization weight μ for a given situation. Therefore, for each object, configuration, SNR level, and regularization, we reconstruct an image for different values of the hyperparameter μ .

The top row of the right panel of Fig. 3 shows the effects of different values of μ , whereas the left panel displays the obtained MSE (see Sect. 3.4) for each values of μ . For too small a value of μ , the under-regularized image (labeled with an A) has plenty of artifacts. In contrast, for too large a μ , the over-regularized image (labeled with a C) is blurred, and many fine features are lost. These two extreme situations correspond to high values of the MSE (A and C), but there is a situation where the MSE reaches a minimum and the image appears to have far fewer artifacts (B).

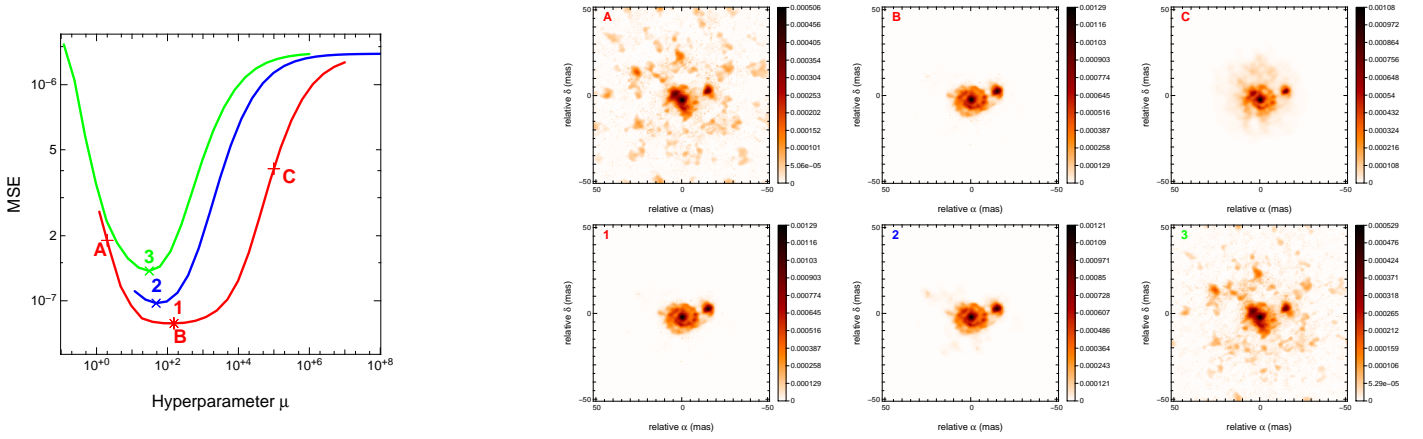


Fig. 3. Left panel shows a plot of the MSE as a function of the hyperparameter μ . The different colors correspond to different levels of SNR (red high, blue intermediate, green poor). For each curve, the optimal value μ^+ is labeled by a number (1, 2, and 3). The corresponding images are shown in the bottom row of the right panel. The top row of the right panel shows three reconstructed images with different values of μ , labeled by a letter on the red curve of the left part (A an under-regularized image, B the best image, and C an over-regularized image). This example is made for the galaxy object and the medium (u, v) coverage. The regularization is the MEM-prior one.

Visual comparison of the A and C images with the one obtained for B confirms that the MSE is a good criterion to correctly set the regularization weight.

We conclude that there is indeed an optimal value of μ , the one that gives the smallest MSE

$$\mu^+ = \arg \min_{\mu} \|\mathbf{x}_{\mu}^{\text{rec}} - \mathbf{x}^{\text{ref}}\|^2, \quad (18)$$

where $\mathbf{x}_{\mu}^{\text{rec}}$ is the image reconstructed with a regularization weight set to μ . As the minimum of the curve is quite flat, the optimal value of μ is not precisely defined but may vary by a factor as large as either two or three with a negligible influence on the result.

This procedure cannot be used in practice because the true image is obviously unknown. Nevertheless, this procedure allows us to define the most accurate image that can be reconstructed given the data and the type of regularization. **In the following analysis, the reconstructed images are always obtained with $\mu = \mu^+$.**

4.2. Dependence of μ^+ on the MSE quality criterion

We wish to determine the type of relationship that links the optimized regularization factor μ^+ to the MSE in order to detect any trends. The scatter plot in Fig. 4 reports the values of μ^+ and MSE obtained for each simulation. In the left panel, the different colors and symbols correspond to different objects. In the right panel, the different colors and symbols correspond to different regularizations. In the bottom row, we present our computed histograms of MSE (left) and μ^+ (right). As in the rest of the paper, the plotted histograms are approximations of the probability density functions (PDF) of our results. These curves were computed from our samples following the optimal method described in Scott (1992).

In the left part of Fig. 4, the different colors representing the objects seem to be aligned vertically. We therefore conclude that the MSE depends mostly on the structure of the observed object.

More precisely, two classes of objects can be distinguished in the distribution of MSE in the left panel of Fig. 4. We therefore grouped together the objects with similar behaviors in the

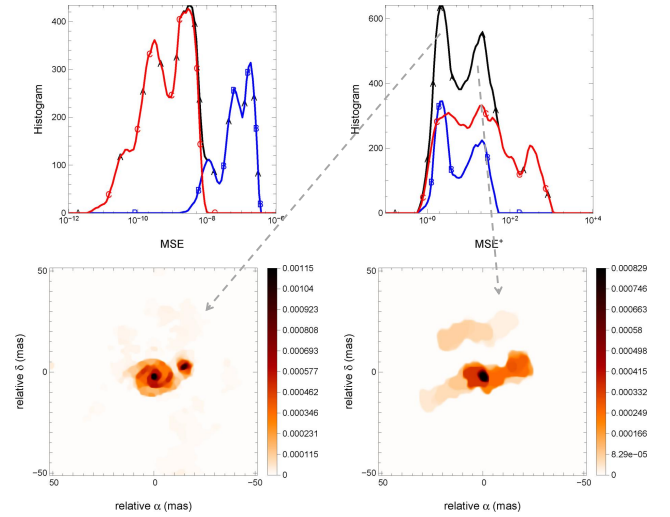


Fig. 5. Upper row: Distribution of the MSE (left) and the MSE⁺ (right). The colors and letters represent the two classes of objects: blue/B for the objects with very compact structures, red/C for the others. The total distribution is shown in the black/A curve. Bottom row: example of reconstructed images for the good (left) and bad (right) MSE⁺ peak.

top panels of Fig. 5: (i) the objects with very compact sources, i.e. the star cluster, the protoplanetary disk and the microlensing (curve in blue labeled B), and (ii) the other objects with extended structures (curve in red labeled C). The MSE is systematically higher for objects of the first class.

Following these observations, we try to find a way of renormalizing the MSE as a function of the object. To join the curves together, we define a new MSE criterion, called MSE⁺, by dividing each MSE by the smallest MSE for each object separately. As expected, this normalization cancels the different object classes, as seen in the right part of Fig. 5. Two peaks clearly appear on the graph, distinguishing the good reconstructions (left peak) from the bad ones (right peak). An example of each case is shown on the bottom part of Fig. 5. The low quality reconstructions are caused by bad configurations (not enough (u, v))

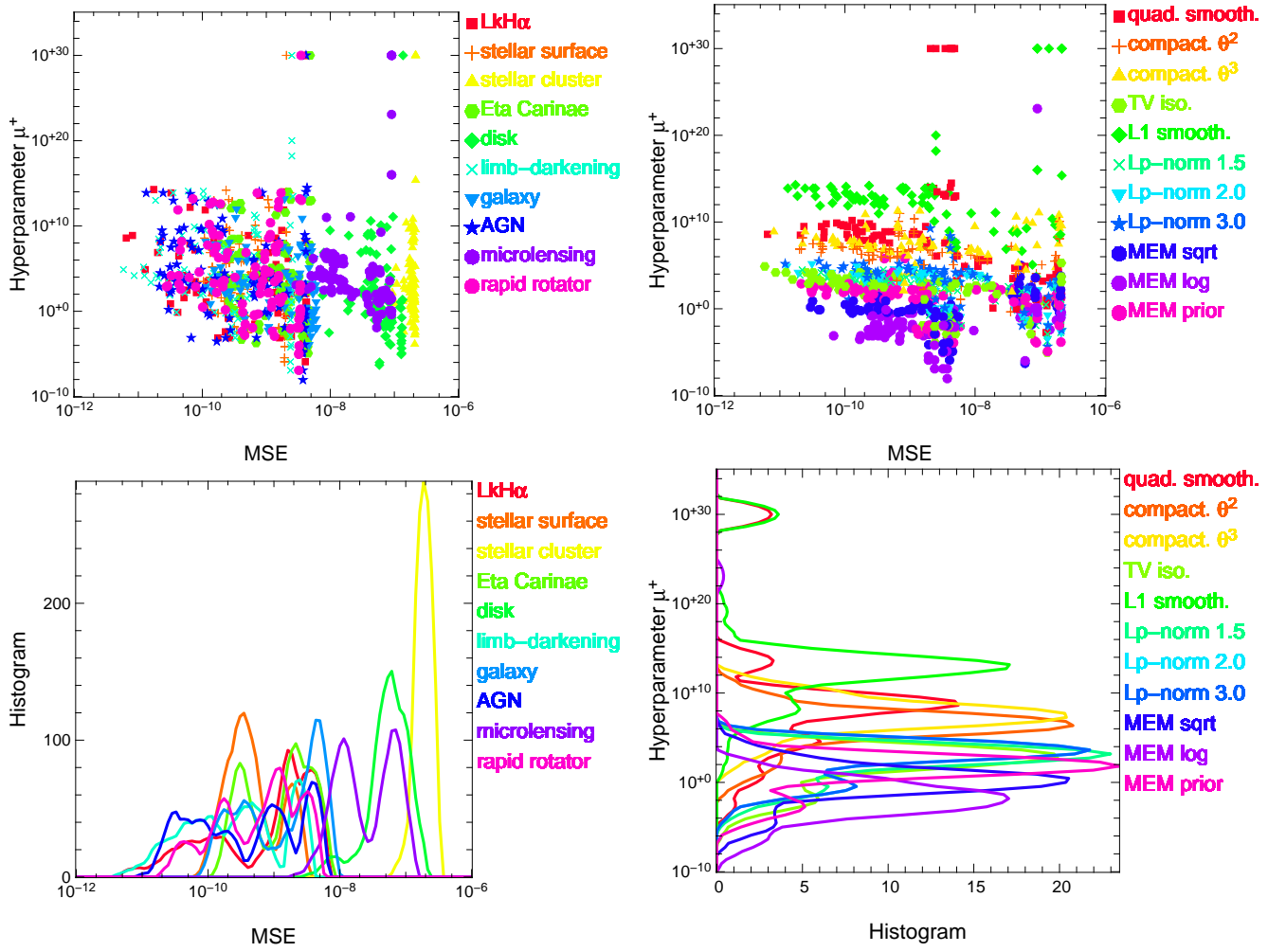


Fig. 4. *Top row:* Scatter plots representing the optimal value of the hyperparameter μ^+ as a function of the MSE of the images. *Bottom row:* The corresponding histograms of MSE and μ^+ . *Left column:* The colors and symbols indicate the different classes of objects. *Right column:* The colors and symbols indicate the different regularization classes.

points, low SNR) and/or bad regularizations as it can be seen in the right part of Fig. 6 and will be described in the next sections. To be sure that the peaks represent the good and bad reconstructions and do not come from different object's types, we verify that each object is represented in each peak, as shown in the left part of Fig. 6.

By visual inspection, we assess that the value of the MSE^+ leads to a correct ordering of the images reconstructed for a given object when the other settings change (data quality, type of regularization, *etc.*): the lower the MSE^+ , the higher the quality of the image is. Other attempts at the renormalization of the MSE are explained in Appendix B.

Now that a good quality criterion has been defined and that the optimal value of μ have been obtained, we can study the other parameters.

4.3. Limits due to the (u, v) coverage and the SNR

In this section, we classify the observational configurations ((u, v) coverage and SNR) on the basis of the MSE^+ . For each data set (unique combination of object and regularization), we order the pair $[(u, v)$ coverage, SNR] according to the value of MSE^+ they reach, giving them a rank from 1 for the best configuration (lowest MSE^+) to 9 for the worst (highest MSE^+). In

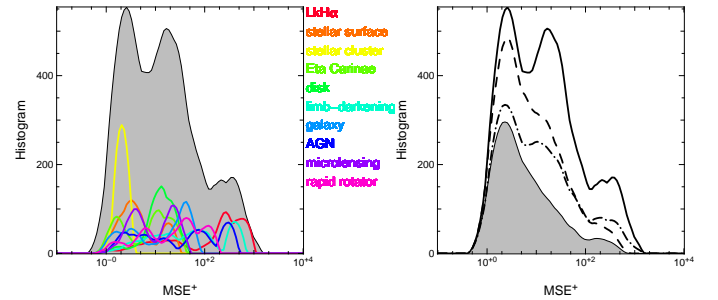


Fig. 6. Distribution of MSE^+ . *Left:* histograms of MSE^+ for different objects in different colors; the gray zone corresponds to the total distribution, all objects confounded. *Right:* solid line, all the configurations and regularizations are kept; dashed line, with the sparsest (u, v) coverage removed; dot-dashed line, with the bad regularizations removed; in gray zone, with the sparsest (u, v) coverage and bad regularizations removed.

the left panel of Fig. 7, we display the cumulative distributions of the ranks reached by every configuration, determining how many times a given configuration reaches at least the first rank, the second rank, *etc.* The highest quality configurations are the ones in the upper-left part of the plot.

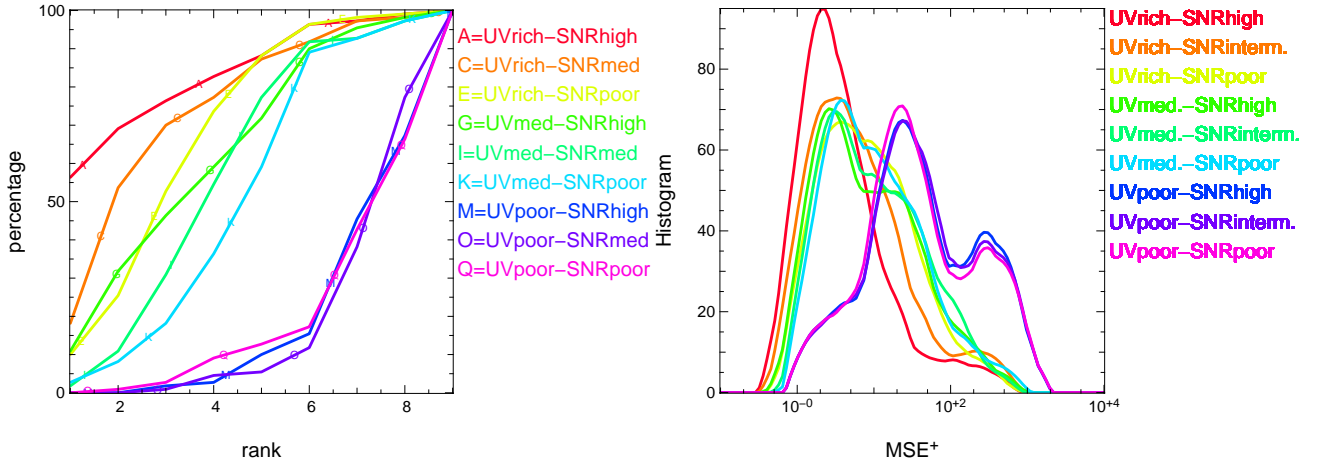


Fig. 7. *Left:* cumulative distributions of the ranks reached by the different configurations of (u, v) coverage and SNR. *Right:* the histograms of the MSE^+ for different configurations of (u, v) coverage and SNR represented by different colors.

The poor (u, v) coverage combined with any value of SNR is clearly too sparse to reconstruct good images. While acceptable reconstruction is possible for any considered SNR when there are enough samples in the (u, v) coverage. We deduce that there is a minimal (u, v) coverage needed to perform image reconstruction, whereas there is no such limit set by the tested SNR. However, when the (u, v) coverage is sufficient, the higher the SNR or the more filled the (u, v) coverage, the higher the quality of the reconstructed image is. The bottom row of the right panel of Fig. 3 shows how the visual quality of the optimal image depends on the SNR. As expected, the higher the SNR, the better the reconstructed image is.

Figure 7 (right) shows the histogram of the MSE^+ for different configurations of (u, v) coverage and SNR. It indicates that the success image reconstruction is more influenced by the amount of data than by the SNR: the MSE^+ is lower for a rich (u, v) coverage with a poor SNR than for a medium (u, v) coverage with an intermediate SNR. We note that all the tested (u, v) coverage are uniform, but we expect that the amount of data has to be sufficient and also *homogeneously spread* in the (u, v) plan.

After the removal of the sparsest coverage, the MSE^+ distribution is shown in Fig. 6 in dashed line. There is still a little bump of bad MSE^+ caused by bad regularizations, as discussed in the next section.

4.4. Quality of the regularizations

Using the same principles as in the previous section, we classify the regularizations as a function of the MSE^+ for different realizations. Figure 8 shows the corresponding cumulative distributions. Isotropic TV appears to be the most successful regularization for the two classes of objects. The compactness prior with $w_n^{\text{prior}} = |\theta_n|^2$ is the second highest quality choice for the very compact objects. The worst regularizations are the ℓ_p -norm, the MEM-sqrt, and the MEM-log, as expected and explained in Sect. 2.3.2. Reconstructions for good and bad regularizations are illustrated in Fig. 9.

In the MSE^+ distribution after the elimination of the bad regularizations (see Fig. 6 in dot-dashed line), there is still much evidence of lower quality MSE^+ . We conclude that the bad MSE^+ are mainly caused by the sparsest (u, v) coverage. However, both have to be eliminated to obtain a cleaned sample of reconstructed images.

In what follows, we made a selection and only retained the regularizations (quadratic and ℓ_1 smoothness, compactness, isotropic TV and MEM with a Gaussian prior) and the (u, v) coverages (rich and medium) that lead to correct reconstructed images. We kept data for all values of SNR.

4.5. Predetermined value of the hyperparameter μ^+ ?

In a fully Bayesian framework, the data noise and the variability of the object are independent. We therefore expect that μf_{prior} does not depend on the data (Tarantola 2005). As a result, for a given type of object, the optimal value for μ should be the same regardless of the SNR or the (u, v) coverage. However, we are not in a truly Bayesian framework because the regularizations are derived from general considerations that must help us to solve for the degeneracies of the problem and are not really derived from the statistics of the brightness distributions of the observed objects. Since the degeneracies of the problem are mostly due to the sparsity of the (u, v) coverage, this observational parameter may have some influence on the tuning of the regularization weight. Our expectation is thus that, for a given type of object, a quasi optimal value of μ^+ can be derived from simulations and from simple considerations to scale this parameter if different image reconstruction settings are used.

The hyperparameter μ^+ depends mostly on the regularizations, as seen in the right part of Fig. 4, where the colors, representing the regularizations, appear aligned horizontally and the pre-eminent peaks of the histograms are separated. Moreover, the hyperparameter μ appears to be quasi independent of the SNR and the (u, v) coverage, as expected. As seen in Fig. 10 for the TV regularization, if there is still a variation in the value of μ , it depends on the object morphology but not on the (u, v) plane and the SNR value.

This finding allows us to link each regularization to a mean value of the hyperparameter μ . This mean value gives a useful start point for the user (see Table 1). The equations to rescale this value in the case of different image settings are given in Appendix C.

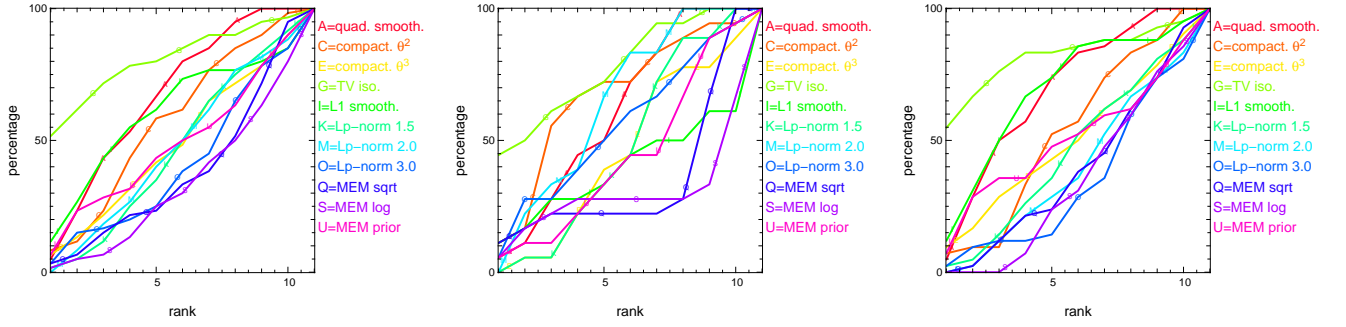


Fig. 8. Cumulative distributions of the ranks reached by the regularizations. *Left:* all objects; *Middle:* objects with very small structures; *Right:* other objects.

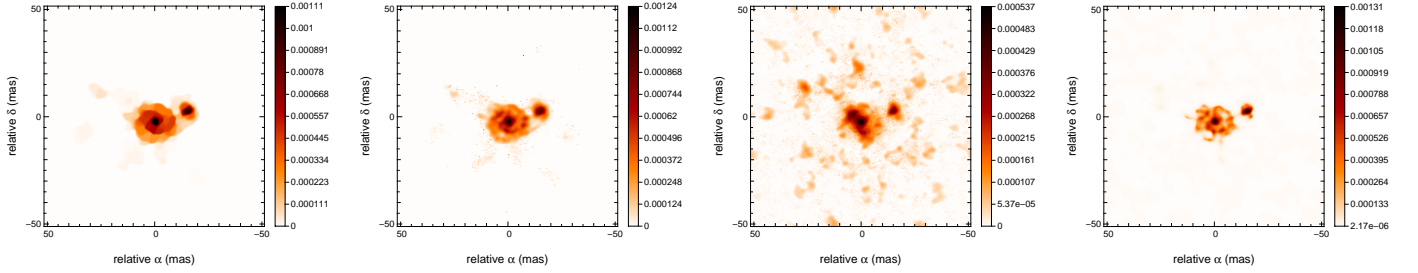


Fig. 9. Example of reconstructed images of the galaxy image with the medium (u, v) coverage and the intermediate SNR for different regularizations. *From left to right,* TV, compactness θ^2 , ℓ_p norm with $p = 2$, MEM-log.

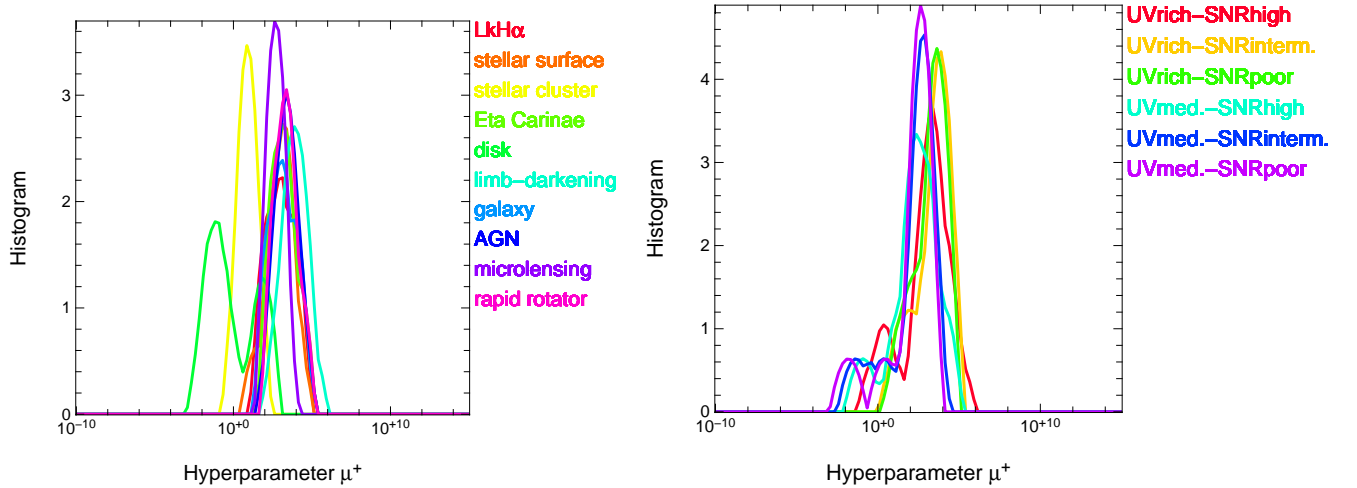


Fig. 10. Histogram of μ^+ for the TV regularization. *Left:* the colors correspond to different objects. *Right:* the colors correspond to different pair $[(u, v)$ coverage–SNR].

Table 1. Table of the mean values of μ for each regularization.

Regularizations	μ^+
Quadratic smoothness	10^9
Compactness r^2	10^7
Compactness r^3	10^7
TV isotropic	10^3
ℓ_1 smoothness	10^{13}
MEM with prior	10^2

4.6. How different noise realizations affect the MSE and the optimal μ ?

In all simulations, in order not to influence the classification of the results, the same random seed is used to compute the noisy

data. Therefore, we test 100 noise realizations for each regularization in the case of the galaxy with a medium (u, v) coverage and an intermediate SNR to study its impact on the curves computed from a single realization. From Fig. 11, we conclude that the MSE is not very different, thus the image reconstruction does not depend on the particular noise realization (as shown for example in Fig. 12). Moreover, the optimal value of μ varies by less than an order of magnitude.

4.7. The effective spatial resolution

To investigate whether super-resolution is achievable and to quantify the amount of trustable super-resolution, we estimate the effective resolution of the reconstructed images. We define the effective resolution as the full width at half maximum (here-

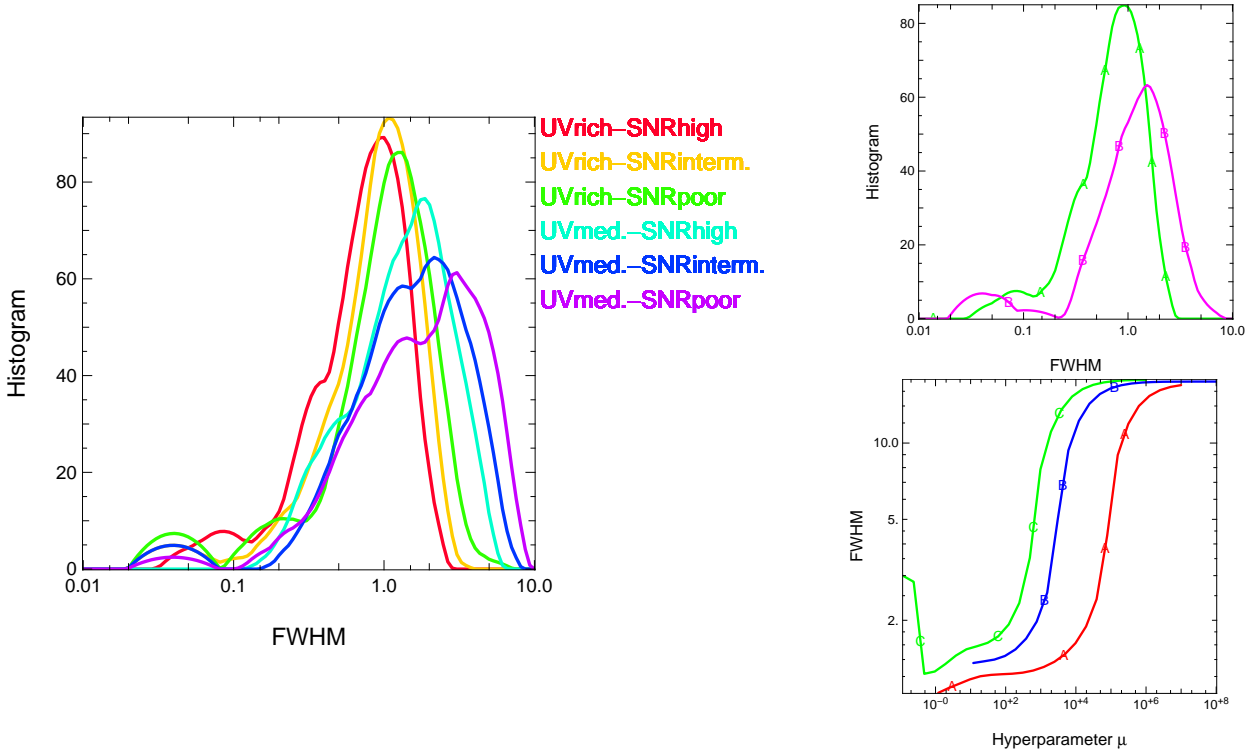


Fig. 13. *Left:* FWHM of the Gaussian computed for the effective resolution in units of the interferometric resolution of the data. *Up-right:* comparison between the FWHM with (green) and without (magenta) the constraint of positivity. *Bottom-right:* variation in the effective resolution as a function of the hyperparameter μ for three different SNR (red high, blue intermediate, green poor). The regularization used is the TV one.

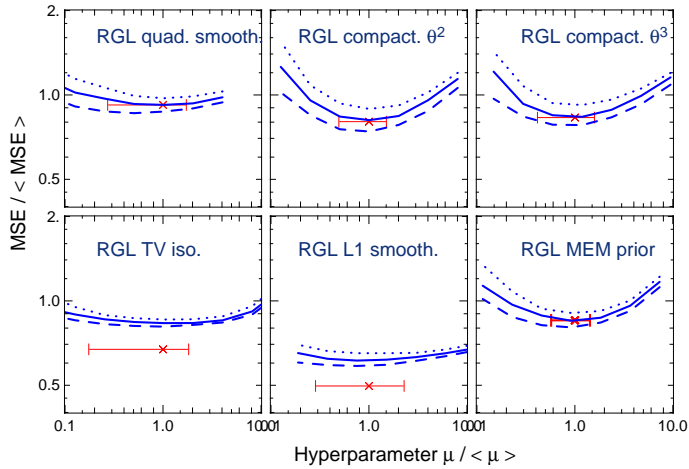


Fig. 11. Variation in MSE and μ with different noise realizations. Blue, the quartile curve of the realizations (25% in dash line, 50% in solid line, and 75% in dot line). Red, the mean optimal μ (cross) and its variance.

after FWHM) of the Gaussian that yields the smallest value of MSE between the reconstructed image and the true image smoothed by that Gaussian

$$\text{FWHM} = \arg \min_w \left\| \mathbf{x}^{\text{rec}} - \mathbf{G}(w) \cdot \mathbf{x}^{\text{ref}} \right\|^2, \quad (19)$$

where $\mathbf{G}(w)$ is a linear smoothing filter that convolves its argument by a Gaussian of FWHM equal to w .

As shown in the left part of Fig. 13, the effective resolution varies with both the amount of data and the SNR: the more ex-

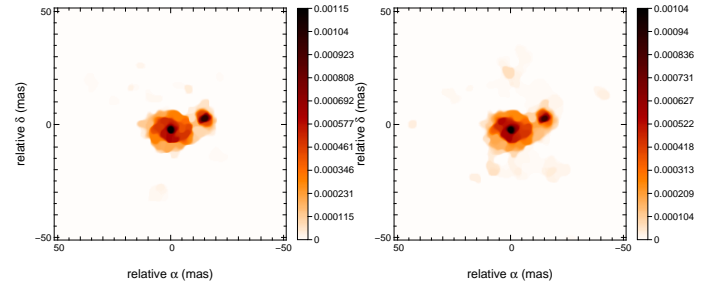


Fig. 12. Reconstructed images for the best (left) and the worst (right) noise realization in the TV case.

tensively the (u, v) parameter space is filled and the higher the SNR, the higher the effective resolution is. As for the MSE (cf. Sect. 4.3), the effective resolution is more influenced by the (u, v) coverage than by the SNR. Super-resolution can be achieved in only the best cases.

The bottom-right part of Fig. 13 shows the expected behavior of the effective resolution: the higher the hyperparameter μ , the lower the effective resolution is. In other words, the more the image is regularized, the fewer the details that are visible.

An interesting question is why this super-resolution exists. Although we do not have a definitive answer, we can speculate that it is due to the role of the positivity in the image reconstruction (Biraud 1969). This is confirmed by our simulations (cf. the upper right part of Fig. 13): without the positivity constraints, the distribution of the FWHMs has a peak higher than when using the positivity constraints.

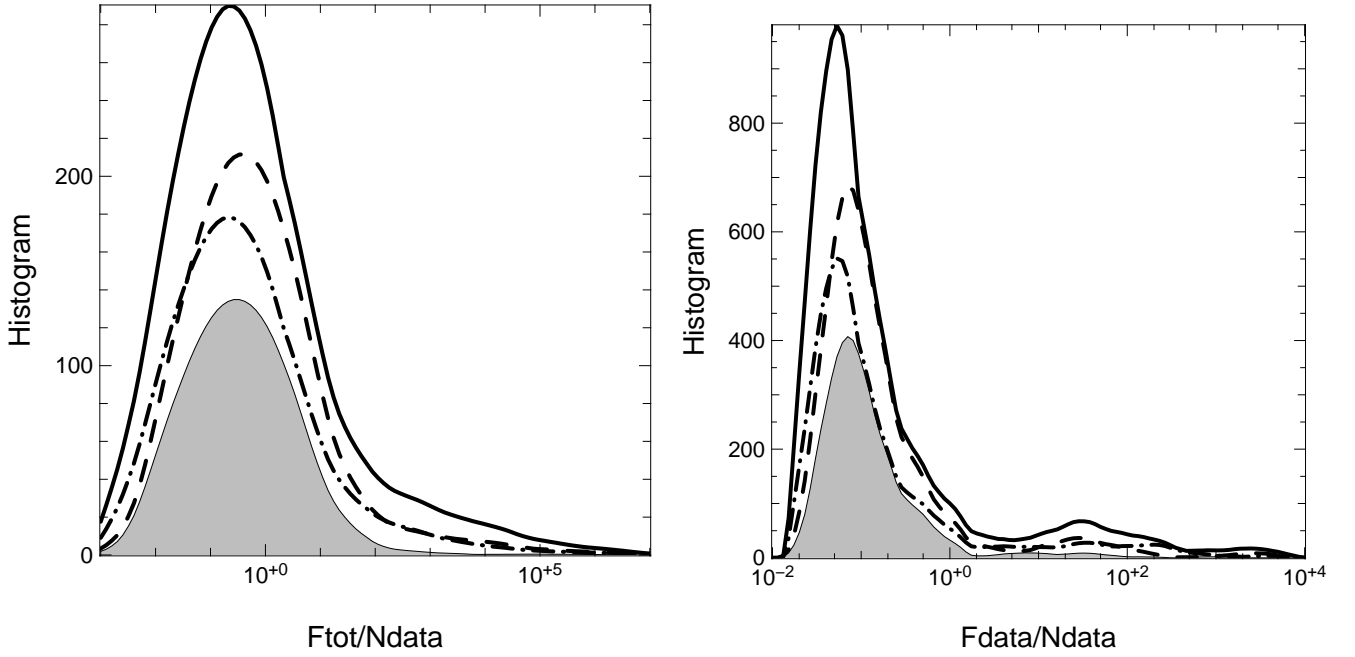


Fig. 14. Histograms of f/N_{data} and $f_{\text{data}}/N_{\text{data}}$. *Solid line*: all configurations and regularizations are kept. *Dashed line*: with the sparsest (u, v) coverage removed. *Dot-dashed line*: with the bad regularizations removed. *In gray zone*: with the sparsest (u, v) coverage and bad regularizations removed.

4.8. Other methods for tuning the regularization?

For fully Gaussian statistics (*i.e.* all the likelihood and prior penalties are quadratic and there are no constraints such as the normalization and the positivity), the expected values of the total penalty function $f(\mathbf{x}^+)$ and the likelihood term $f_{\text{data}}(\mathbf{x}^+)$ at the optimal solution \mathbf{x}^+ is given by (Tarantola 2005)

$$\mathbb{E}\{f(\mathbf{x}^+)\} = N_{\text{data}}, \quad (20)$$

$$\mathbb{E}\{f_{\text{data}}(\mathbf{x}^+)\} = N_{\text{data}} - N_{\text{edf}}, \quad (21)$$

where N_{data} is the number of data points (for both visibility amplitudes and phases), and N_{edf} is the number of equivalent degrees of freedom. Despite our being unable to use fully Gaussian statistics, the left panel of Fig. 14 shows that the distribution of $f(\mathbf{x}^+)$ peaks approximatively at the value of N_{data} . The spread of this distribution however prevents us to be able to tune the regularization level according to the criterion that $f(\mathbf{x}^+) \approx N_{\text{data}}$.

The value of $f_{\text{data}}(\mathbf{x}^+)/N_{\text{data}}$ is around 0.1, which means that $N_{\text{edf}}/N_{\text{data}} \approx 90\%$ of the data information is resolved. The image reconstruction is able to estimate almost as many parameters as data points. Since the $f_{\text{data}}(\mathbf{x}^+)/N_{\text{data}}$ ratio has a smaller range (from ~ 0.3 to ~ 0.003) than the possible value of the hyperparameter μ (from 100 to 10^{13}), it may be easier to tune the balance between the terms in the penalty function thanks to the $f_{\text{data}}(\mathbf{x}^+)/N_{\text{data}}$ value instead of μ . However, this criterion may be really variable with the noise statistics and a study of the variation in $f_{\text{data}}(\mathbf{x}^+)/N_{\text{data}}$ with different noise statistics should be done before using it as a tune factor (Gull 1988; Pichon & Thiebaut 1998).

Another way of determining the value of μ is the so-called L-curve. The L-curve is a log-log plot of the regularization term f_{prior} versus the likelihood term f_{data} for a range of the hyperparameter μ . In the L-curve criterion, the regularization parameter μ is such that the corresponding point on the L-curve lies in the corner. This choice is motivated by the corner being the separation between the flat part where the solution is dominated by

regularization errors and the vertical part where it is dominated by the perturbation errors (Hansen 2000). The correct behavior of the L-curve is confirmed by the simulations, as seen in Fig. 15 (right): since the curves are plotted as a function of the highest quality image, they should cross in their corner, which is globally the case. We note that the shape of the L-curve seems more complicated as there are at least two corners and not only one (see Fig. 15 left). The L-curve appears to be an appropriate tool for finding the optimal value of μ but a more general study has to be done to confirm its suitability (see comparison with GCV) and its practical implementation in an image reconstruction algorithm.

5. Conclusions

Thanks to the use of a flexible algorithm devoted to image reconstruction in optical interferometry, we have performed a detailed study of the regularization terms. This study is the first one to compare such a number of regularizations on an equal footing, *i.e.* with the same algorithm and using the same data type. Performing these systematic tests has allowed us to discuss the different parameters and terms in the image reconstruction and to extract some practical rules, which are summarized in the following:

1. A **minimal (u, v) coverage** is necessary to reconstruct an image. Even if the image quality improves with the SNR, such a limit does not exist for the SNR. In other words, in the image reconstruction technique, increasing the number of telescopes is more interesting than constructing larger ones. The homogeneity of the (u, v) coverage is probably also critical but has not been tested.
2. Some regularizations are suitable for the optical image reconstruction and others not, regardless of the object being targeted. The holes in the (u, v) plane are the major issue in optical interferometry and the main role of the regularization is the correct interpolation of the missing data, regardless

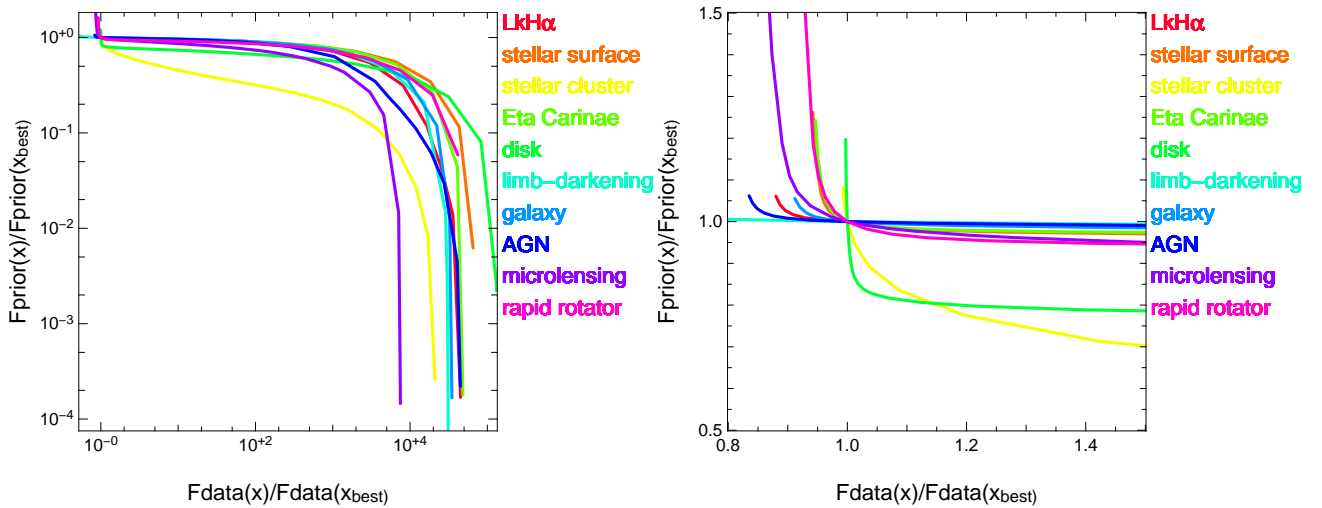


Fig. 15. An example of the L-curve in the simulations for the TV regularization with a zoom on the right part.

of the object. The highest quality regularization among the tested ones is the **isotropic total variation**.

3. The **hyperparameter** μ does not depend on the (u, v) coverage and the SNR as theoretically expected and depends mostly on the type of regularization. An optimal value for each regularization tested in this paper is given in Table 1. This optimal value may vary by a factor of 2-3 without there being any major changes in the images. A slight dependence on both the object structures and pixel size is also discernible and the equation to rescale the optimal values are computed. It should be interesting to implement regularizations independent of the pixel size.
4. **Super-resolution** can be achieved in the image reconstruction process and its level rises with the (u, v) coverage filling.
5. There are various possible ways of tuning the regularization level:
 - The visual tuning is enough as the μ value can slightly vary without causing any large changes in the reconstructed image.
 - Setting the likelihood term f_{data} seems to be a more effective way of fixing the balance between the regularization and the likelihood terms. However, the variation in the likelihood term with noise statistics needs to be investigated.
 - The L-curve criterion could give correct results.

Acknowledgements. This research has made use of Yorick, a free data processing language written by D. Munro (<http://yorick.sourceforge.net>) and the Service de Calcul Intensif de l’Observatoire de Grenoble.

References

- Biraud, Y. 1969, *A&A*, 1, 124
 Candes, E. J., Romberg, J., & Tao, T. 2006, *IEEE Transactions on Information Theory*, 52, 489
 Charbonnier, P., Blanc-Fraud, L., Aubert, G., & Barlaud, M. 1997, *IEEE Trans. Image Process.*, 6, 298
 Delpiancke, F., Deric, F., Paresce, F., et al. 2003, *Ap&SS*, 286, 99
 Domiciano de Souza, A., Vakili, F., Jankov, S., Janot-Pacheco, E., & Abe, L. 2002, *A&A*, 393, 345
 Filho, M. E., Renard, S., Garcia, P., et al. 2008, in Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference, Vol. 7013, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series
 Goodman, J. W. 1985, *Statistical Optics* (John Wiley & Sons)
 Gull, S. F. 1988, in *Maximum Entropy and Bayesian Methods in science and engineering*, ed. G. J. Erickson and C. R. Smith

- Gull, S. F. & Skilling, J. 1984, in *Indirect Imaging. Measurement and Processing for Indirect Imaging*, ed. J. A. Roberts, 267–+
 Hansen, P. C. 2000, in *Computational Inverse Problems in Electrocardiology*, ed. P. Johnston, *Advances in Computational Bioengineering* (WIT Press), 119–142
 Haubois, X., Perrin, G., Lacour, S., et al. 2009, *A&A*, 508, 923
 Hestroffer, D. 1997, *A&A*, 327, 199
 Kraus, S., Hofmann, K., Menten, K. M., et al. 2010, *Nature*, 466, 339
 Kraus, S., Weigelt, G., Balega, Y. Y., et al. 2009, *A&A*, 497, 195
 Lawson, P. R., ed. 2000, *Principles of Long Baseline Stellar Interferometry*
 Lawson, P. R., Cotton, W. D., Hummel, C. A., et al. 2004, in *Bulletin of the American Astronomical Society*, Vol. 36, *Bulletin of the American Astronomical Society*, 1605–+
 Le Besnerais, G., Lacour, S., Mugnier, L. M., et al. 2008, *IEEE Journal of Selected Topics in Signal Processing*, Vol. 2, Issue 5, p.767–780, 2, 767
 Le Bouquin, J., Lacour, S., Renard, S., et al. 2009, *A&A*, 496, L1
 Malbet, F. & Perrin, G. 2007, *New Astronomy Review*, 51, 563
 Meimon, S., Mugnier, L. M., & Le Besnerais, G. 2005, *Journal of the Optical Society of America A*, 22, 2348
 Monnier, J. D., Zhao, M., Pedretti, E., et al. 2007, *Science*, 317, 342
 Narayan, R. & Nityananda, R. 1986, *ARA&A*, 24, 127
 Pauls, T. A., Young, J. S., Cotton, W. D., & Monnier, J. D. 2005, *PASP*, 117, 1255
 Pearson, T. J. & Readhead, A. C. S. 1984, *ARA&A*, 22, 97
 Pichon, C. & Thiebaud, E. 1998, *MNRAS*, 301, 419
 Renard, S., Malbet, F., Benisty, M., Thiebaud, E., & Berger, J. 2010, *A&A*, 519, A26+
 Scott, D. W. 1992, *Multivariate density estimation: theory, practice, and visualization* (John Wiley & Sons, Inc.)
 Strong, D. & Chan, T. 2003, *Inverse Problems*, 19, S165
 Tarantola, A. 2005, *Inverse Problem Theory and Methods for Model Parameter Estimation* (SIAM)
 Thiebaud, E. 2005, in *NATO ASIB Proc. 198: Optics in astrophysics*, ed. R. Foy & F. C. Foy, 397–+
 Thiebaud, E. 2008, in Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference, Vol. 7013, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series
 Thiebaud, E. & Giovannelli, J.-F. 2010, *IEEE Signal Processing Magazine*, 27, 97
 Titterton, D. M. 1985, *A&A*, 144, 381
 Zhao, M., Gies, D., Monnier, J. D., et al. 2008, *ApJ*, 684, L95
 Zhao, M., Monnier, J. D., Pedretti, E., et al. 2009, *ApJ*, 701, 209

Appendix A: The regularization expressions

In our simulations, we test 11 different regularization terms commonly used in image reconstruction methods and implemented in MiRA:

1. Quadratic smoothness:

$$f_{\text{prior}}(\mathbf{x}) = \|\mathbf{x} - \mathbf{S} \cdot \mathbf{x}\|^2, \quad (\text{A.1})$$

where \mathbf{S} is a smoothing operator implemented via finite differences.

2-3. Compactness (Le Besnerais et al. 2008):

$$f_{\text{prior}}(\mathbf{x}) = \sum_n w_n^{\text{prior}} x_n^2, \quad (\text{A.2})$$

which is a separable quadratic regularization. To enforce compactness, the weights $w_n^{\text{prior}} > 0$ have to increase with the distance from the center of the image. Two different cases were studied in the simulations: $w_n^{\text{prior}} = \|\theta_n/\Delta\theta\|^2$ and $w_n^{\text{prior}} = \|\theta_n/\Delta\theta\|^3$, where $\|\theta_n/\Delta\theta\|$ is the distance in pixels of the n -th pixel from the center of the field of view.

4. Total variation (Strong & Chan 2003):

$$f_{\text{prior}}(\mathbf{x}) = \sum_{n_1, n_2} \sqrt{\|\nabla x_{n_1, n_2}\|^2 + \epsilon^2}, \quad (\text{A.3})$$

where

$$\|\nabla x_{n_1, n_2}\|^2 = (x_{n_1+1, n_2} - x_{n_1, n_2})^2 + (x_{n_1, n_2+1} - x_{n_1, n_2})^2$$

is the squared magnitude of the spatial gradient in the image, $\epsilon > 0$ is a small number inserted to avoid the discontinuity in zero, and $(n_1, n_2) \sim n$ are the two dimensional indices of the n -th pixel. In our reconstructions, ϵ has always been chosen so as to be negligible compared to the gradient of significant structures.

5. $\ell_2 - \ell_1$ smoothness (Charbonnier et al. 1997):

$$f_{\text{prior}}(\mathbf{x}) = \tau^2 \sum_n \psi(\|(\mathbf{D} \cdot \mathbf{x})_n\|/\tau), \quad (\text{A.4})$$

where $\psi(z) = z - \log(1 + z)$ is a $\ell_2 - \ell_1$ norm, $\tau > 0$ is a threshold level, and \mathbf{D} is a finite difference operator approximating the q -th spatial derivatives of its argument. When $\|(\mathbf{D} \cdot \mathbf{x})_n\|$ is much smaller than the threshold, $\tau^2 \psi(\|(\mathbf{D} \cdot \mathbf{x})_n\|/\tau) \approx 1/2 \|(\mathbf{D} \cdot \mathbf{x})_n\|^2$, while, for $\|(\mathbf{D} \cdot \mathbf{x})_n\|$ it is much larger than the threshold, $\tau^2 \psi(\|(\mathbf{D} \cdot \mathbf{x})_n\|/\tau) \approx \tau \|(\mathbf{D} \cdot \mathbf{x})_n\|$. This regularization attempts to strongly smooth the weak spatial gradients and slightly smooth the strong gradients. In our simulations, we take \mathbf{D} to approximate the spatial Laplacian and choose a threshold small enough such that the regularization behaved mostly like a linear smoothness.

6-8. ℓ_p -norm:

$$f_{\text{prior}}(\mathbf{x}) = \sum_n (x_n^2 + \epsilon^2)^{p/2} \approx \sum_n |x_n|^p, \quad (\text{A.5})$$

where $\epsilon > 0$ is a small value introduced to avoid the singularity in zero when $p \leq 1$. For $p > 1$, the ℓ_p -norm regularization tends to produce a smooth image as it reduces the variance of the pixels. For $p < 1$, the ℓ_p -norm regularization tends to promote sparsity in the image. This is interesting mostly for objects consisting of point sources. True sparsity constraints would be obtained for $p = 0$, although when $p < 1$ the regularization is no longer convex and the optimization problem becomes extremely difficult to solve as p gets closer to 0. Results in compress sensing (Candes et al. 2006) have proven that choosing $p = 1$ yields the most sparse solution, like $p = 0$ for a large class of problems. However, taking $p = 1$ yields a non-smooth but convex problem that is much easier to solve than the combinatorial problem resulting from

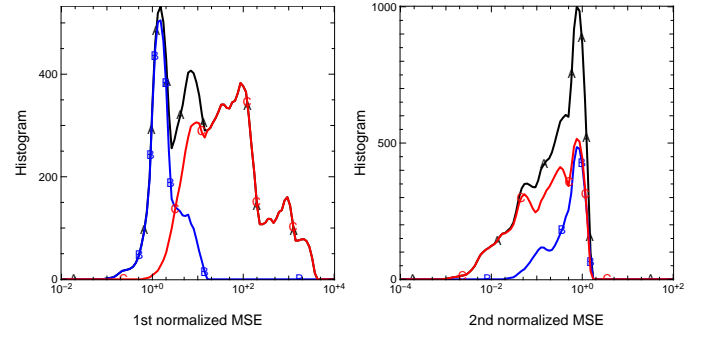


Fig. B.1. Attempt to renormalize the MSE. Distribution of the first normalized MSE (left) and of the second normalized MSE (right). The colors and letters represent the two classes of objects: blue/B for the objects with very compact structures, red/C for the others. The total distribution is shown in black/A curve.

the choice $p = 0$. With positivity and normalization constraints, the ℓ_1 -norm of \mathbf{x} is constant. Hence, taking $p = 1$ is meaningless in our framework and we consider only $p = 1.5$, $p = 2$ and $p = 3$.

9-11. Maximum entropy methods (Narayan & Nityananda 1986):

$$f_{\text{prior}}(\mathbf{x}) = - \sum_n h(x_n; \bar{x}_n). \quad (\text{A.6})$$

Here the prior is to assume that the image is drawn towards a prior model $\bar{\mathbf{x}}$ according to a non-quadratic potential h , called the *entropy*. We try three entropies:

$$\text{MEM-sqrt: } h(x; \bar{x}) = \sqrt{x}; \quad (\text{A.7})$$

$$\text{MEM-log: } h(x; \bar{x}) = \log(x); \quad (\text{A.8})$$

$$\text{MEM-prior: } h(x; \bar{x}) = x - \bar{x} - x \log(x/\bar{x}). \quad (\text{A.9})$$

MEM was first introduced in radioastronomy and is useful for images made of bright point-like sources on a smooth background. In our simulations, we took the prior image $\bar{\mathbf{x}}$ to be the isotropic 2-D Gaussian that most accurately fits the amplitude visibility data.

Appendix B: Renormalization of MSE

We attempted to normalize the MSE with two additional methods:

1. Since the squared difference between the real image and a smoothed version of the real image is higher for images with sharp or point-like structures, we computed a first *normalized* MSE as

$$\text{MSE}_{\text{norm},1} = \frac{\sum_n (x_n^{\text{rec}} - x_n^{\text{ref}})^2}{\sum_n ((\mathbf{S} \cdot \mathbf{x}^{\text{ref}})_n - x_n^{\text{ref}})^2}, \quad (\text{B.1})$$

where \mathbf{S} is a smoothing operator. The distribution of this normalized MSE is shown in the right panel of Fig. B.1: the distribution is narrower but the two classes remain, despite the normalization.

2. In the second normalization, the MSE is compared to the norm of the reference image

$$\text{MSE}_{\text{norm},2} = \frac{\sum_n (x_n^{\text{rec}} - x_n^{\text{ref}})^2}{\sum_n (x_n^{\text{ref}})^2}. \quad (\text{B.2})$$

This normalization is more effective than the previous one, as it joins the curves together. However, the distribution is unimodal and does not enable us to distinguish the good and the bad reconstructions. It is thus not a useful normalization.

Appendix C: Scaling the hyperparameter μ

In a Bayesian framework, the prior penalty $\mu f_{\text{prior}}(\mathbf{x})$ should only depend on both the sought after brightness distribution $I(\boldsymbol{\theta})$ and the image parametrization. In this appendix and from these simple principles, we derive a method to adapt the value of the hyperparameter μ when the image parameters such as the pixel size are modified.

Using the sampled image model in Eq. (4), the normalization of \mathbf{x} implies that

$$1 = \sum_n x_n = \alpha \sum_n I(\boldsymbol{\theta}_n) \approx \frac{\alpha}{(\Delta\theta)^L} \int I(\boldsymbol{\theta}) d^L\boldsymbol{\theta},$$

where we have used the Riemann approximation of integrals, $\Delta\theta$ is the pixel size, and $L = 2$ is the number of dimensions in $\boldsymbol{\theta}$. Without any loss of generality, we can assume that the sought after brightness distribution is normalized such that $\int I(\boldsymbol{\theta}) d^L\boldsymbol{\theta} = 1$; hence

$$\alpha \approx (\Delta\theta)^L, \quad (\text{C.1})$$

and Eq. (4) becomes

$$x_n \approx (\Delta\theta)^L I(\boldsymbol{\theta}_n). \quad (\text{C.2})$$

C.1. Separable ℓ_p norm

Using Eq. (4) and then, the Riemann approximation, the prior penalty for a separable ℓ_p norm regularization, after substituting in Eq. (A.5), becomes

$$\begin{aligned} \mu f_{\text{prior}}(\mathbf{x}) &= \mu \sum_n |x_n|^p \approx \mu \alpha^p \sum_n |I(\boldsymbol{\theta}_n)|^p \\ &\approx \mu \frac{\alpha^p}{\Delta\theta^L} \int |I(\boldsymbol{\theta})|^p d\boldsymbol{\theta}, \end{aligned} \quad (\text{C.3})$$

which shows that for a regularization by a separable ℓ_p norm

$$\mu \propto \begin{cases} \Delta\theta^L / \alpha^p & \text{in general;} \\ \Delta\theta^{L(1-p)} & \text{with the normalization constraint.} \end{cases} \quad (\text{C.4})$$

Hence, with the normalization constraint, the optimal value of μ should be the same regardless of the pixel size for a regularization given by a separable ℓ_1 norm.

C.2. ℓ_p norm on the gradient

Using 1-D notation to simplify the equations, the prior penalty for a regularization by the ℓ_p norm on the gradient is given by

$$\begin{aligned} \mu f_{\text{prior}}(\mathbf{x}) &= \mu \sum_n |x_{n+1} - x_n|^p \\ &\approx \mu \alpha^p \sum_n |I(\theta_n + \Delta\theta) - I(\theta_n)|^p \\ &\approx \mu \alpha^p \Delta\theta^p \sum_n |\partial_\theta I(\theta_n)|^p, \end{aligned}$$

where $\partial_\theta I(\boldsymbol{\theta})$ denotes the partial derivative of the brightness distribution along the angular direction. In L dimensions and using the Riemann approximation, this gives

$$\mu f_{\text{prior}}(\mathbf{x}) \approx \mu \alpha^p \Delta\theta^{p-L} \int |\partial_\theta I(\boldsymbol{\theta})|^p d\boldsymbol{\theta},$$

which shows that

$$\mu \propto \begin{cases} \Delta\theta^{L-p} / \alpha^p & \text{in general;} \\ \Delta\theta^{L-p(L+1)} & \text{with the normalization constraint.} \end{cases} \quad (\text{C.5})$$

Applying this result for a regularization by quadratic smoothness in 2D, *e.g.* Eq. (A.1), we found that, with a normalization constraint, $\mu \propto \Delta\theta^{-4}$.

C.3. Total variation

The preceding result, with $p = 1$, readily applies to regularization by the total variation, that is

$$\mu \propto \begin{cases} \Delta\theta^{L-1} / \alpha & \text{in general;} \\ \Delta\theta^{-1} & \text{with the normalization constraint.} \end{cases} \quad (\text{C.6})$$

We can also deduce that, if a relaxed version of TV is used, as in Eq. (A.3), the relaxation parameter ϵ must scale as the pixel size $\Delta\theta$ to have the prior penalty approximately insensitive to the pixel size.

We note that, with our particular choice of the threshold τ for the $\ell_2 - \ell_1$ smoothness regularization defined in Eq. (A.4), we expect this regularization to behave mostly like TV.

C.4. Quadratic compactness

The quadratic compactness we used in MiRA is given by Eq. (A.2)

$$\begin{aligned} \mu f_{\text{prior}}(\mathbf{x}) &= \mu \sum_n \left\| \frac{\boldsymbol{\theta}_n}{\Delta\theta} \right\|^q x_n^2 \approx \mu \frac{\alpha^2}{\Delta\theta^q} \sum_n \|\boldsymbol{\theta}_n\|^q I(\boldsymbol{\theta}_n)^2 \\ &\approx \frac{\mu \alpha^2}{\Delta\theta^{q+L}} \int \|\boldsymbol{\theta}\|^q I(\boldsymbol{\theta})^2 d\boldsymbol{\theta}, \end{aligned}$$

with $q = 2$ or 3 . From this last approximation, we derive the scaling of μ with the pixel size

$$\mu \propto \begin{cases} \Delta\theta^{q+L} / \alpha^2 & \text{in general;} \\ \Delta\theta^{q-L} & \text{with the normalization constraint.} \end{cases} \quad (\text{C.7})$$

Hence, in 2D ($L = 2$) and for a normalized image, μ does not depend on the pixel size for $q = 2$, while it scales as $\Delta\theta$ for $q = 3$.

C.5. Maximum entropy

For maximum entropy methods, we have

$$\mu \sum_n \sqrt{x_n} \approx \mu \frac{\alpha^{1/2}}{\Delta\theta^L} \int I(\boldsymbol{\theta})^{1/2} d\boldsymbol{\theta},$$

and

$$\mu \sum_n h(x_n; \bar{x}_n) \approx \mu \frac{\alpha}{\Delta\theta^L} \int h(I(\boldsymbol{\theta}); \bar{I}(\boldsymbol{\theta})) d\boldsymbol{\theta},$$

with $h(x; \bar{x}) = x - \bar{x} - x \log(x/\bar{x})$ as in Eq. (A.9). from which we deduce that

$$\mu \propto \begin{cases} \Delta\theta^L/\alpha^{1/2} & \text{for MEM-sqrt;} \\ \Delta\theta^L/\alpha & \text{for MEM-prior.} \end{cases} \quad (\text{C.8})$$

Hence, for a normalized image, μ does not depend on the pixel size in a MEM-prior regularization.